

Weekly Report (2014.04.21~04.27)

Done

- 1) 这周大部分时间还是花在了集群上，把上周末完成的 **hadoop + hdfs** 扫了个尾，又接着布了 **Hbase**、**Hive**，然后和小马一起把所有的数据都拷贝到集群上。不幸的是拷贝完数据网线貌似没接好，节点都连不上，周末就啥都干不了了。
通过集群部署，还是学到很多新的知识的，之前只是停留在简单地使用 **MapReduce** 本身，对 **Hbase**、**Hive** 等 **Hadoop** 生态圈内的项目都知之甚少，通过这次这个项目可以有不错的补充。
- 2) 七巧板部分，从“高多维变量”过滤到“低多维变量”的工作还是需要补上，和解聪进行了简短的讨论，觉得相对熵不太适合，互信息、转移熵、皮尔森系数可以用。
变量过滤之后的后续分析部分，张老师建议看看变分贝叶斯(**variational bayesian network**)，看了一些其他文献，还有时序布尔网络、微分方程等方法。我会先去了解一下，再和两位老师讨论。

To Do

- 1) 现在数据已经在集群上了，但是没有导入数据库，这个可能要花不少时间，我们对 **Hive** 或者 **shark** 都没有经验，这些数据库相比于传统的数据库需要配置的东西更多。
- 2) 万事开头难。。。尽量在气象项目上有点实质性的进展。